International Journal of
Applied Mathematics,
Electronics and Computers

ISSN: 2147-8228 http://ijamec. atscience.org

Advanced Technology and Science

Original Research Paper

Improvement of Solution Problems of Matrix Equations Calculated by Computers#

Levent Civcik*[1], Cevdet Çetin[2], Haydar Bulgak[3]

Accepted 15th August 2014 **DOI: 10.18100/ijamec.87886**

Abstract: Solution of linear algebra systems may come out with "ill-condition" or "well-condition" based on input information and solution methods. The aim of this study is to determine and correction of problems that may come out from the solution of matrix equations by computers and to calculate $A.x = f$ linear algebra.

Keywords: *The Linear Algebra System, Condition Number, Epsilon.*

1. Introduction

If we solve $A.x = f$ equation system computer, we have to know about numbers subset in the numbering system that called "FORMAT".
It is

$$F(\gamma, P\_, P\_+, K) = \{0\} \cup \{z : z = \pm \gamma^{P(z)} m(z),$$
$$m(z) = a_1 / \gamma + a_2 / \gamma^2 + ... + a_K / \gamma^k$$
$$a_1 \neq 0, \quad 0 \le a_j \le \gamma - 1, \qquad j = 1, 2, ..., k$$
$$P\_ \le P(z) \le P\_+$$

If we take $\gamma, P\_, P\_+$ as constants than we can accept

$$F_k = (\gamma, P\_, P\_+, k)$$

In this set there is

$$\varepsilon_\infty = \gamma^{P\_+}(1 - \gamma^{1-k})$$

$-\varepsilon_\infty$ $+\varepsilon_\infty$

The boundaries $[-\varepsilon_\infty, +\varepsilon_\infty]$ chance for any computer depending on the coprocessor in the computer.
In general as a rule there are two computer constants that characterize the approach. We show these constants by ε_1 and ε_0 symbols. ε_1 is used for approaching to $"1"$, ε_0 is approaching to $"0"$ in computer. Practically there is no number in the ranges $(0, \varepsilon_0)$ and $(1, 1+\varepsilon_1)$ in computer.

A. Error in Placing The Real Number into Format

Let's take real numbers set of R and set F define an operator

$$fl : [-\varepsilon_\infty, +\varepsilon_\infty] \cap R \to F$$

on these two sets, so that is places real number into format. If $z \in F$ then $fl(z) = z$. Otherwise we can't determine any number into format that shows z, if $z > \varepsilon_\infty$ or $z < -\varepsilon_\infty$ that is fl operator is shown in different form in the computer. But all of operator have some common points.
We have to point out that there will be "overflow error" when we put numbers beyond the interval $[-\varepsilon_\infty, +\varepsilon_\infty] \cap R$. So we don't define fl value for these numbers. But computer user has to take some precaution to avoid these problems.
Now let's give the upper boundary of the error that come out when we place number Z into the format. It is clear that the value of error after fl operation is:

$$|z - fl(z)| = \alpha |z| + \beta$$

Here $|\alpha| < \varepsilon_1, |\beta| < \varepsilon_0$, and $\alpha.\beta = 0$. The value $|z - fl(z)|$ after fl operation is called Round off error. $fl(z)$ can be written $|z - fl(z)| = \alpha |z| + \beta$.
When we place number Z into the computer we see three causes.

1. If z is the element of format, it will be place into the memory without any error.
2. If instead of z, we place $fl(z)$, it will be placed into the memory as the closest number to the z value. Then

$$|z - fl(z)| \le \varepsilon_0$$

or

$$|z - fl(z)| \le \varepsilon_1 |z|$$

error come out the total error will be shown as

$$|z - fl(z)| \le \varepsilon_1 |z| + \varepsilon_0$$

[1]*Higher School of Vocational and Technical Sciences Department of Computer Technologies, Selcuk University, Campus, Konya/Turkey*
** Corresponding Author: Email: lcivcik@selcuk.edu.tr*
 # This paper has been presented at the International Conference on Advanced Technology&Sciences (ICAT'14) held in Antalya (Turkey), August 12-15, 2014.

If z is out of chosen format then computer will show *overflow error a*nd will stop the operation.

Similarly, A is a real matrix in the NxN type and is real vector of $Nx1$ type. If there will be no error in the placing to the computer, providing A and f big enough

$$(\|A\|) > \varepsilon_0 \text{ and } \|f\| > \varepsilon_0)$$
$$\|A - fl(A)\| \le \varepsilon_1 \|A\|$$
$$\|f - fl(f)\| \le \varepsilon_1 \|f\|$$

condition provide. This is called Input error. Here is the spectral norm of matrix A , and is the Euclid norm of vector f . This means,

$$\|f\| = \sqrt{\sum_{i=1}^{N} f_i^2} \quad ; \qquad \|A\| = \max_{\|x\|=1} \|Ax\|$$

B. Error That may Come Out When We Solve $A.x = f$ Equation in The Computer

Now, let's see the main problem. What can we do to calculate linear equation with two diagonals in computer? Let's take the following system:

$$a_1 x_1 + b_2 x_2 = f_1$$
$$+a_2 x_2 + b_3 x_3 = f_1$$
$$.....$$
$$a_{N-1} x_{N-1} + b_N x_N = f_{N-1}$$
$$a_N x_N = f_N$$

Here a_i , b_i and f_i are given real number, x_i is the element of desired vector $(i = 1, 2, ..., N$, $j = 2, 3, ..., N)$.

$$x_N = f_N / a_N$$
$$x_{N-1} = (f_{N-1} - b_N x_N) / a_{N-1}$$
$$.....$$
$$x_1 = (f_1 - b_2 x_2) / a_1$$

This is necessary and if we apply this as an example in the computer.

$$x_1 - 2x_2 = 0$$
$$x_2 - 2x_3 = 0$$
$$.....$$
$$x_{N-1} - 2x_N = 0$$
$$x_N = 1$$

Here, all diagonal elements are equal to 1; i.e., according to previous rule, solution of the problem is present and unique.

$$x_j = 2^{N-j} \quad , \quad j = 1, 2, ..., N$$

But, this is a risky solution for computer side. Because, set of numbers in computer is restricted and there no any number bigger than ε_∞ .

So, for any ε_∞ , on condition of $2^{N-j} > \varepsilon_\infty$, there is a number $N = N(\varepsilon_\infty)$. If, N is selected like this way, we meet to an overflow error when we solve given problem at computer.

In classical mathematics education, the well-known Cramer rule is used for solving $A.x = f$ equation, it is based on $\det A$. We saw before that is unsuccessful. In addition, let us give another example that is shown $\det A$ is unsuccessful.

Let we take following equations system:

$$1/2 x_1 = 1$$
$$1/2 x_2 = 1$$
$$.....$$
$$1/2 x_N = 1$$

In this example, because of $a_j = 1/2$, $j = 1, 2, ..., N$, the solution is present and unique. The co-efficient matrix of this system is diagonal matrix. It is clear that, let's take this:

$$\det A = 1/2^N$$

Therefore, for any ε_∞ on condition of $2^{N-1} > \varepsilon_\infty$, there is a number $N = N(\varepsilon_\infty)$ .That is, the computer assume that $1/2^N$ is zero. According to that, $\det A$ became zero, $\det A = 0$. Thus, the absence of a well-condition problem's solution is given as a result.

We can say that at problem solving with computers some methods given in classical mathematics, like Cramer rule, cause to obtain unsuccessful results.

2. Examination The Problem with $\mu(A)$ Condition Number

In computer-aided mathematical education, if it was started to solve a problem by using some methods that no having robust basics, it is clear that we will have important problems. This has been appearing in the last 50 years. Some studies have been done about this subject by Neuman [4] and Turing [3]. After that, for example, studies given by [1] [2] and for the defined problem at above the condition number is

specified $\mu(A) = \|A\| \|A^{-1}\|$ and all problems have dealed with this $\mu(A)$.

The condition number has two important attributes. We will see these at following theorems.

Theorem 1: if $\mu(A) < \infty$ and $2\mu(A)\|B\| / \|A\| < 1$ then, in this case

$$\mu(A+B) < \infty \text{ and } |\mu(A) - \mu(A+B)| \le 3\mu^2(A)\|B\| / \|A\|$$

Theorem 2: $\mu(A) < \infty$ and $2\mu(A)\|B\| / \|A\| < 1$, in this case the nearness of and nearest of $Ax = f$ end nearest $(A+B)y = f + g$ solutions is given by following inequality:

$$\|x - y\| / \|x\| \le 3\mu(A)(\|B\| / \|A\| + \|g\| + \|f\|)$$

The proof of these theorems could be found at [1].

It is clear that if the $\mu(A)$ is given bigger, the solution of $Ax = f$ is effected more from changing A and f . That is, if smaller, the problem is well-condition.

In applications, the correctness of data specifies two different ways. 1. Approximation and reading errors. 2. Representation and truncation errors for data stored in computer memory. Therefore naturally instead of $Ax = f$ equation, we accept the $(A+B)y = f + g$ equation the "neighborhood" to it. We know

that only $\|B\|/\|A\| < eps$ and $\|g\|/\|f\| < eps$. Here, the eps parameter specifies that input data accuracy. So, the eps parameter is very important. If errors are only representation errors, in this case eps= ε_1.

If we solve the $Ax = f$ problem, we don't need the more accreted system in practice. Because, we can't explain about the computed solution is near the real solution. In this situation, the solution is to use a parameter that accepted the upper limit of practical invertible matrices [5,6]. By using this parameter, we can express the Theorem 1 in other way.

Theorem 3: If $\mu(A) < \mu^*$ and $2\mu^*\|B\|/\|A\| < 1$ then

$$\mu(A+B) < 1.5\mu^* \text{ and}$$

$$|\mu(A) - \mu(A+B)| \le \mu^2(A)\|B\|/\|A\|$$

This theorem is related how much correct the value of μ^* and input values given. According to this theorem, $\mu^* < \|A\|/(2\|B\|)$ must be held.

If $20\mu(A)\varepsilon_1 < 1$ is held, for the problem $A.x = f$, we can find an exact solution in the computer like $\|x - y\| < \varepsilon_1\|x\|$

Thus, it can be defined by computer as $\mu^* = 1/(20\varepsilon_1)$ naturally.

It is important to know that what kind of problem we are given, namely whether it is a good one or bad one. There are many different criterions depending on the type of the problem. For example, the problem of $A.x = f$, well-known in mathematics, has always a solution for nonzero f. As long as $\det A \neq 0$ holds. The solution exists and it is unique. If $\det A \neq 0$ does not hold, there are still some approaches to solve the problem. However, the solution either non-exists or it is not unique. That means this specific problem is bad for us.

$$A.x = f \quad , \quad \det A \neq 0$$

For the problem above, there exists a solution and its unique independently from f (for each f). This is a well-defined problem. But, as soon as the elements of A are changed, the state of the problem also changes. In other words, it may not be a well defined problem anymore. For instance, for

$$A = \begin{pmatrix} \varepsilon & 0 \\ 0 & 1 \end{pmatrix}$$

$\det A = \varepsilon \neq 0$ and for each of f, there exists as solution and it is unique. On the other hand,

$$B = \begin{pmatrix} -\varepsilon & 0 \\ 0 & 0 \end{pmatrix}$$

For B above, the problem $(A+B)y = f$ have infinite solution for some f while some have no solution since $\det(A+B) = 0$. The existence and uniqueness of the solution depend on the elements of f.

In such problems, it can be thought that, how far or how near the problem is to the bad-defined region. How much we can change the elements of A, so it is still in a well-defined region.

For the problem $(A+B)y = f$, if $A.x = f$ is in a well conditioned region, as long as $(\|B\|/\|A\|) < (1/10\mu(A))$ holds for a small normed B, the problem $(A+B)y = f$ is also in a well defined region. Here, $\mu(A) = \|A\|\|A^{-1}\| \ge 1$ is condition number for the equation $A.x = f$. Hence, for

$$A = \begin{pmatrix} \varepsilon & 0 \\ 0 & 1 \end{pmatrix}$$

while $A.x = f$ is well-defined, in order to be $(A+B)y = f$ well defined, since $\mu(A) = 1/\varepsilon$. $\|A\| = 1$, for $B\|B\| < \varepsilon/10$ must hold. In other words for all Bs, which hold the inequality $\|B\| < \varepsilon/10(A+B)$ is well defined.

Such problems whose solutions exist, are unique and stable in terms of their elements are called well-conditioned (correct).

The set of matrices can be divided into two parts which are the subset of well- conditioned ones and the subset of ill-conditioned ones by the condition number $\mu(A)$, we can illustrate this situation by a Fig.1:

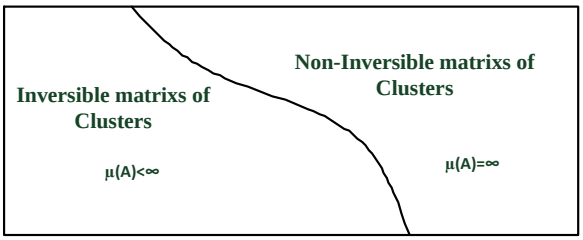

Figure 1

If $\mu(A) < \infty$ holds then the problem $A.x = f$ is a well-conditioned problem. It is possible to say that the distance of a matrix A to the region of ill-conditioned matrices is $1/[10\mu(A)]$. Lets it by Fig.2:

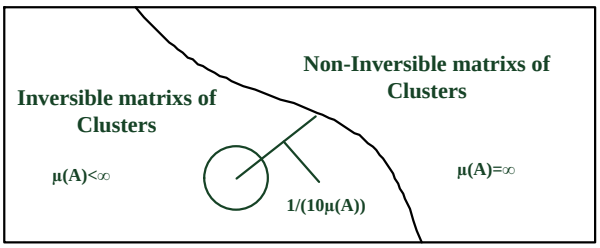

Figure 2

There are many studies concerning with the distance of a matrix to the region for well-conditioned matrices.

In most of these studies indefiniteness principle arises. Namely, when we get a matrix A from the set $\{A, \|A-D\| < \varepsilon\|D\|, \text{ is the given matrix, } \mu(D) < \infty\}$ if $\mu(D)\varepsilon > 1$ holds, we can not still say that whether matrix A is in the set of well-conditioned ones or not. Although matrix D is in the set of well-conditioned, it is too close the set of ill-conditioned ones. It can be illustrated by a Fig 3.

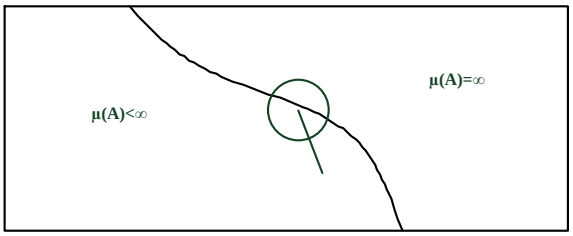

Figure 3

It should be defined that practical well-conditioned matrices so that we can obtain the distance of a matrix to the set of the ill-conditioned matrices.

As long as we are given a number μ^* which holds $\mu(A) = \|A\|\|A^{-1}\| < \mu^* < \infty$, we can say that the problem $A.x = f$ is a well-conditioned problem practically. Otherwise it is a ill-conditioned one. Here the number μ^* is a boundary between practically well- conditioned problems and practically ill-conditioned problems. We can illustrate this situation by a Fig.4.

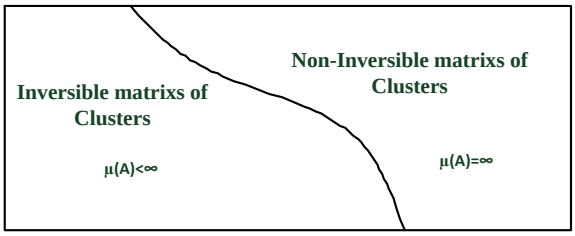

Figure 4

3. Conclution

So, the most important advantage of this approach is to warn the users about the problem (system) whether it is a well-conditioned or ill-conditioned problem by computing the exact value of $\mu(A)$ by given formula; then if necessary, they change the data and input values which make the problem ill-conditioned or improve the approach partially or fully so that they have a chance to prevent themselves from wasting their time and work.

Acknowledgment

This study is based from master's thesis which name is "Matris Problemlerine Bilgisayar Destekli Bazı Çözümler" by Levent Civcik.

References

[1] S.K. Godunov, "Solving Linear Algebraic Systems", Novosibirsk, Nauka, (in Russian), 1980.

[2] J.H.Wilkinson, "The Algebraic Problem. Clarendom Press", Oxford, 1965.

[3] A.M. Turing, "Rounding-off errors in matrix processes", Quart. J. Mech., 1, 287-308, 1948.

[4] J. von Neuman and H.H. Goldstine, "Numerical inverting of matrices of high cjjrder", Bull. Amer. Math. Soc., v.53, no. 11, 1021-1099, 1947.

[5] L. Civcik, "Matris Problemlerine Bilgisayar Destekli Bazı Çözümler", Selçuk University, Konya, 1998.

[6] A. Bulgak, "Cheking of a well-conditioning of the interval matrices", Sibearian J. of Differential Equations, N.-Y., Nova Science Publ. (to appear).